

Caching at the Edge: a Green Perspective for 5G Networks

Bhanukiran Perabathini^{*,†}, Ejder Baştug[◇], Marios Kountouris^{*,◇}, Mérouane Debbah[◇] and Alberto Conte[†]

[◇]Large Networks and Systems Group (LANEAS), CentraleSupélec, 91192, Gif-sur-Yvette, France

^{*}Department of Telecommunications, CentraleSupélec, 91192, Gif-sur-Yvette, France

[†]Alcatel Lucent Bell Labs, 91620, Nozay, France

{ejder.bastug, marios.kountouris, merouane.debbah}@supelec.fr,
{bhanukiran.perabathini, alberto.conte}@alcatel-lucent.com

Abstract—Endowed with context-awareness and proactive capabilities, caching users' content locally at the edge of the network is able to cope with increasing data traffic demand in 5G wireless networks. In this work, we focus on the energy consumption aspects of cache-enabled wireless cellular networks, specifically in terms of area power consumption (APC) and energy efficiency (EE). We assume that both base stations (BSs) and mobile users are distributed according to homogeneous Poisson point processes (PPPs) and we introduce a detailed power model that takes into account caching. We study the conditions under which the area power consumption is minimized with respect to BS transmit power, while ensuring a certain quality of service (QoS) in terms of coverage probability. Furthermore, we provide the optimal BS transmit power that maximizes the area spectral efficiency per unit total power spent. The main takeaway of this paper is that caching seems to be an energy efficient solution.

Index Terms—5G, mobile wireless networks, caching, energy efficiency, area power consumption, stochastic geometry

I. INTRODUCTION

Fueled by the ubiquity of new wireless devices and smart phones, as well as the proliferation of bandwidth-intensive applications, user demand for wireless data traffic has been growing tremendously, increasing the corresponding network load in an exponential manner. This is further exacerbated by rich media applications associated with video streaming and social networking. In parallel, both academia and industry are now in the urge of evolving traditional cellular networks towards the next-generation broadband mobile networks, coined as 5G networks, targeting to satisfy the mobile data tsunami while minimizing expenditures and energy consumption. Among these intensive efforts, caching users' content locally at the edge of the network is considered as one of the most disruptive paradigms in 5G networks [1].

Interestingly yet not surprising, recent results have shown that distributed content caching can significantly offload different parts of the network, such as in radio access networks (RANs) and core network (CN), by smartly prefetching and storing content closer to the end-users [2]. Indeed, traditional cellular networks, which have been designed for mobile devices with limited processing and storage capabilities, have

started incorporating context-aware and proactive capabilities, fueled by recent advancements in processing power and storage. As a result, caching has recently taken the 5G literature by storm.

The idea of caching at the edge of the network, namely at the level of base stations (BSs) and user terminals (UTs), have been highlighted in various works, including edge caching [3], FemtoCaching [4], and proactive caching [5]. Spatially random distributed cache-enabled BSs are modeled in [6], and expressions for the outage probability and average delivery rate are derived therein. Another stochastic framework but in the context of cache-enabled device-to-device (D2D) communications is presented in [7], studying performance metrics that quantify the local and global fraction of served content requests. Although most prior work (see [8] for a recent survey) deals with different aspects of caching, i.e. performance characterization and approximate algorithms, the energy consumption behavior of cache-enabled BSs in densely deployed scenarios has not been investigated. This is precisely the focus of this paper.

Energy consumption aspects of cellular networks are usually investigated by placing BSs on a regular hexagonal or grid topology and conducting intensive system-level simulations [9]. Therein, the aim is to find the optimal values of parameters, such as cell range and BS transmit power, while minimizing the total power consumed under certain quality-of-service (QoS) constraints. Despite its attractiveness, it is rather cumbersome or even impossible to analytically evaluate key performance metrics in large-scale networks. A common simplification in modeling cellular networks is to place the BSs on a two dimensional plane according to a homogeneous spatial Poisson point process (PPP), enabling to handle the problem analytically [10]. Several works studied the validity of PPP modeling of BSs compared to regular cellular models (e.g. [11]) and additional insights can be obtained by analytically characterizing performance metrics, such as the spatial distribution of signal-to-interference-plus-noise ratio (SINR), coverage probability, and average rate [12], [13].

In this work, we analyze energy consumption aspects of cache-enabled wireless network deployments using a spatial model based on stochastic geometry, which to the best of our knowledge has not been addressed in prior works. We consider

This research has been supported by the ERC Starting Grant 305123 MORE (Advanced Mathematical Tools for Complex Network Engineering), the projects 4GinVitro and BESTCOM.

the problem of area power consumption (APC) and energy efficiency (EE) of cache-enabled BSs in a scenario where BSs and mobile users are distributed according to independent homogeneous PPPs. We provide conditions under which the area power consumption is minimized subject to a certain quality of service (QoS) in terms of coverage probability. Furthermore, we provide the optimal BS transmit power that maximizes the energy efficiency, defined as the ratio of area spectral efficiency over the total power consumed. The main result of this paper is that distributed content caching in wireless cellular networks turns out to be an energy efficient solution.

II. SYSTEM MODEL

We consider the downlink scenario of a single-tier cellular network in which the BSs are distributed on the two-dimensional Euclidean plane \mathbb{R}^2 according to a homogeneous spatial PPP with density λ_b denoted by $\Phi_b = \{r_i\}_{i \in \mathbb{N}}$, where $r_i \in \mathbb{R}^2$ is the location of the i -th BS. The mobile UTs (or users) are modeled to be distributed on the same plane according to an independent homogeneous PPP with density $\lambda_u > \lambda_b$, which is denoted by $\Phi_u = \{s_j\}_{j \in \mathbb{N}}$ with $s_j \in \mathbb{R}^2$ being the location of the j -th user. Without loss of generality, we focus on a *typical* UT placed at the origin of the coordinate system for calculating the key performance metrics of interest.

In this system model, we assume that the signal transmitted from a given BS is subject to two propagation phenomena before reaching a user: (i) a distance dependent pathloss governed by the pathloss function $g(r) = br^{-\alpha}$, where b is the pathloss coefficient and α is the pathloss exponent, and (ii) Rayleigh fading with mean 1. Therefore, the signal strength from the i -th BS as received by the typical user can be expressed as

$$p_i(r_i) = h_i P b r_i^{-\alpha}, \quad (1)$$

where the random variable h_i denotes the power of Rayleigh fading and P is the transmitted power. In addition, we assume that background noise is present in the system, with variance $\sigma^2 = \beta \lambda_b$, with $\beta = B \frac{1}{\lambda_u} \frac{F k T}{b}$, where B is the total available bandwidth, F is the receiver noise figure, k is Boltzmann constant, and T is the ambient temperature.

As alluded earlier, the typical user connected to the i -th BS receives a signal of power $p_i(r_i)$. This in turn means that the sum of the received powers from the rest of the BSs contributes to the interference to this signal. As a result, the received SINR at the typical user is given by

$$\text{SINR} = \frac{h_i g(r_i) P}{\sigma^2 + I_i}, \quad (2)$$

where $I_i = \sum_{r_j \in \Phi_b \setminus r_i} p_j(r_j)$ is the cumulative interference experienced from all the BSs except the i -th BS.

In a downlink scenario, although the typical user may technically be served from any BS, a connection with a particular BS has to be established according to a association policy that satisfies a certain performance metric, e.g. ensuring QoS. In this work, we assume that a mobile user connects to

the BS that provides the maximum SINR. This is formally expressed as

$$\max_{r_i \in \Phi_b} \text{SINR}(i) > \gamma,$$

where γ is the target SINR. Given the above definition, the typical user is said to be *covered* when there is at least one BS that offers an SINR $> \gamma$. If not, we say that the typical user is not covered. We assume that $\gamma > 1$, which is needed to ensure that there is at maximum one BS that provides the highest SINR for a user at a given instant [14].

A. Coverage and Quality of Service

Among the main objectives of wireless networks is to guarantee a certain QoS for the users. The choice of a specific QoS metric influences both the complexity and the operation of the network. In this paper, we consider the coverage probability as the QoS that has to be assured. The coverage probability for strongest BS association for a general pathloss function $g(r)$ is given as [14], [15]

$$\begin{aligned} P_{\text{cov}}(P, \lambda_b) &= \mathbb{P}[\text{SINR} \geq \gamma] \\ &= \pi \lambda_b \int_0^\infty \exp(-q(P, \lambda_b, r)) dr, \end{aligned} \quad (3)$$

where

$$q(P, \lambda_b, r) = \frac{\gamma \sigma^2}{P g(\sqrt{r})} + \lambda_b \int_0^\infty \frac{\pi \gamma g(\sqrt{r_i})}{g(\sqrt{r}) + \gamma g(\sqrt{r_i})} dr_i. \quad (4)$$

Denoting by $P_{\text{cov}}^{\text{NN}}$ the coverage probability in the case of no noise ($\sigma^2 \rightarrow 0$), the expression for the coverage probability in (3) may be approximated in a low noise regime as [16]

$$P_{\text{cov}}(P, \lambda_b) = P_{\text{cov}}^{\text{NN}} \left(1 - \frac{A'}{\lambda_b^{\frac{\alpha}{2}-1} P} \right), \quad (5)$$

where $A' = \frac{\beta \Gamma(1+\frac{\alpha}{2})}{b[C(\alpha)]^{\frac{\alpha}{2}}}$ and $C(\alpha) = \frac{2\pi^2}{\alpha} \csc\left(\frac{2\pi}{\alpha}\right)$.

The specific QoS constraint we consider here is that the coverage probability experienced by a typical user has to be always greater than the coverage probability achieved in the absence of noise, i.e.

$$P_{\text{cov}} \geq P_{\text{cov}}^{\text{NN}}. \quad (6)$$

As a consequence, it can be shown (from (5) and (6)) that the optimal BS density λ_b^* that guarantees the above QoS constraint is

$$\lambda_b^* \geq \frac{A}{P^{\frac{2}{\alpha-2}}}, \quad (7)$$

with $A = A'^{\frac{2}{\alpha-2}}$. We use this expression in the optimization problem we formulate in Section III.

Once the typical user is covered with a certain SINR, the traffic requests of this typical user have to be satisfied from its BS, by bringing the content from its source on the Internet via the backhaul. In practice, even though the typical user is covered and benefits from high SINR, it is obvious that any kind of bottleneck in the backhaul may result into long delays to the content, degrading the overall quality of experience (QoE). As part of dealing with this bottleneck, we assume

that the BSs are able to store the users' (popular) content in their caches, so that requests can be satisfied locally, without passing over the limited backhaul. This is detailed in the following section.

B. Cache-enabled Base Stations

Several studies have shown that multiple users actually access the same content very frequently. Take for instance some popular TV shows, the case of viral videos with over a billion viewings, news blogs, online streaming, etc. In this context, the network will be inundated with requests for the same content that might largely increase the latency or, eventually, congest the network itself. Otherwise stated, certain types of content (or information) are relatively more popular than others and are requested/accessed more often by the users [17]. Therefore, it is reasonable to assume that a user's choice distribution matches with the global content popularity distribution. As mentioned before, the logic behind having cache-enabled BSs is to exploit this likelihood and store locally at the BSs serving a typical user the content with highest demand (popular demand) so that both users and service providers get an incentive when a popular request is made.

Let us assume that each BS is equipped with a storage unit (hard disk) which caches popular content. Since the storage capacity cannot be infinite, we assume that at each BS a set of content up to f_0 (the catalog) is stored on the hard disk. Rather than caching uniformly at random, a smarter approach will be to store the most popular content according to the given global content popularity statistics. We model the content popularity distribution at a typical user to be a right continuous and monotonically decreasing probability distribution function (PDF), denoted as [18]

$$f_{\text{pop}}(f, \eta) = \begin{cases} (\eta - 1) f^{-\eta}, & f \geq 1, \\ 0, & f < 1, \end{cases} \quad (8)$$

where f is a point in the support of the corresponding content, and η represents the steepness of the popularity distribution curve. We define the steepness factor to be the (average) number of users per BS, that is $\eta = \frac{\lambda_u}{\lambda_b}$. The justification for the above model is that the higher the number of users attached to a BS, the more accurately the trend is sampled, hence the more content is sorted towards the left of the distribution thereby making it steeper. Moreover, since $\lambda_u > \lambda_b$, we have that $\eta > 1$.

Now, given the fact that BSs cache the catalog according to the content popularity distribution in (8), the probability that a content demanded by a connected user falls within the range $[0, f_0]$ is given by

$$\begin{aligned} \mathbb{P}_{\text{hit}} &= \int_0^{f_0} f_{\text{pop}}(f, \eta) df \\ &= \int_0^{f_0} (\eta - 1) f^{-\eta} df \\ &= 1 - f_0^{1-\eta}. \end{aligned} \quad (9)$$

It can be verified that \mathbb{P}_{hit} converges to 1 when $f_0 \rightarrow \infty$, namely when the catalog stored in the BSs goes to infinity. Consequently, the probability that a request is missing from the catalog can be expressed as $\mathbb{P}_{\text{miss}} = f_0^{1-\eta}$. An illustration of the system model is given in Fig. 1, including snapshots of PPPs and visualization of the content popularity distribution. In the following, we introduce a power model which takes into account the caching capabilities at BSs, and will be used for investigating the energy aspects of cache-enabled BS deployment.

C. Power Consumption Model

In the context of studying caching at the edge, there is a large scope for adopting an extensive power model that takes into account various detailed factors. In this paper, we deliberately restrict ourselves to a basic power model as a first attempt to relate energy efficiency to a cache enabled wireless network. We address a more complete model in an extended version of this paper. We consider two different power models depending on whether BSs have caching capabilities or not.

1) *With caching*: The components of the total power consumed at an operating BS is given as follows:

- i) A constant transmit power, P .
- ii) An operational charge at each BS, P_o .
- iii) Power needed to retrieve data from the local hard disk when a content from the catalog is requested, P_{hd} .
- iv) Power needed to retrieve data from the backhaul when a content outside the catalog is requested, P_{bh} .

We assume that $P_{\text{bh}} > P_{\text{hd}}$ motivated by the realistic constraint that it is more power consuming to utilize the backhaul connection than to retrieve stored information from the local caching/storage entity.

Therefore, the total power consumed at a given BS is sum of all the components such as

$$\begin{aligned} P_{\text{tot}}^{(c)} &= P + P_o + P_{\text{bh}} \times \mathbb{P}_{\text{miss}} + P_{\text{hd}} \times \mathbb{P}_{\text{hit}} \\ &= P + P_o + P_{\text{hd}} + (P_{\text{bh}} - P_{\text{hd}}) f_0^{1-\eta} \\ &= P + P_s + P_d f_0^{1-\eta}, \end{aligned} \quad (10)$$

where $P_s = P_o + P_{\text{hd}}$ and $P_d = P_{\text{bh}} - P_{\text{hd}}$.

2) *Without caching*: In the absence of caching, the BS has to retrieve the requested content from the backhaul every service timeslot. This is equivalent to the case where $\mathbb{P}_{\text{hit}} = 0$ (or $\mathbb{P}_{\text{miss}} = 1$). The total power consumed at a given BS is given as

$$\begin{aligned} P_{\text{tot}} &= P + P_o + P_{\text{bh}} \\ &= P + P_s + P_d. \end{aligned} \quad (11)$$

III. AREA POWER CONSUMPTION

The power expenditure per unit area, also termed as APC, is an important metric to characterize the deployment and operating costs of BSs, also indicating the compatibility of the system with the legal regulations. In our system model, the APC for cache-enabled BSs is defined as

$$\mathcal{P}^{(c)} = \lambda_b P_{\text{tot}}^{(c)}. \quad (12)$$

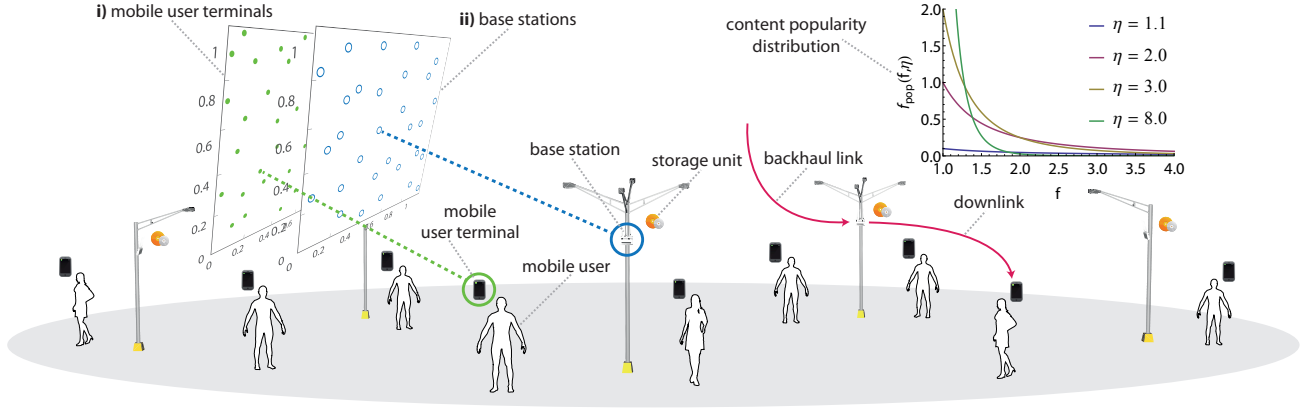


Figure 1: An illustration of the considered system model. The snapshots of PPPs for i) mobile users and ii) base stations on unit area are given on the left side. The content popularity distribution for different values of η is given on the top right side, showing that lower values of η corresponds to a more uniform behavior.

In the same way, the APC of BSs with no caching capabilities is defined as $\mathcal{P} = \lambda_b P_{\text{tot}}$. In this section, we aim at minimizing separately the APC for both cases (caching and no caching), while satisfying a certain QoS. This can be formally written as

$$\begin{aligned} & \underset{P \in [0, \infty)}{\text{minimize}} && \mathcal{P}(P) \text{ or } \mathcal{P}^{(c)}(P) \\ & \text{subject to} && P_{\text{cov}}(P, \lambda_b) \geq P_{\text{cov}}^{\text{NN}}. \end{aligned} \quad (13)$$

Consider first the case where BSs have caching capabilities. The following result can be obtained for the solution of the optimization problem in (13).

Proposition 1. *Suppose that BSs have caching capabilities, thus $\mathcal{P}^{(c)}(P)$ is the objective (utility) function in (13). Then, for $\alpha = 4 + \epsilon$, $\epsilon > 0$, the optimal power allocation P^* that solves (13) is lower bounded as*

$$P^* > \frac{2P_s}{\epsilon}. \quad (14)$$

Proof: Using the expression for optimum λ_b^* from (7) and incorporating it in (12), we get the expression for APC as

$$\mathcal{P}^{(c)} = \frac{A}{P^{\frac{2}{\alpha-2}}} (P + P_s + P_d f_0^{1-\frac{\lambda_u}{A} P^{\frac{2}{\alpha-2}}}). \quad (15)$$

Let ϵ be a real number and write $\alpha = 4 + \epsilon$.

$$\begin{aligned} \mathcal{P}^{(c)} &= \frac{A}{P^{\frac{2}{2+\epsilon}}} (P + P_s + P_d f_0^{1-\frac{\lambda_u}{A} P^{\frac{2}{2+\epsilon}}}) \\ &= A(P^{1-\frac{2}{2+\epsilon}} + P_s P^{-\frac{2}{2+\epsilon}} + P_d P^{-\frac{2}{2+\epsilon}} f_0^{1-\frac{\lambda_u}{A} P^{\frac{2}{2+\epsilon}}}). \end{aligned} \quad (16)$$

For $\epsilon \leq 0$, $\mathcal{P}^{(c)}$ is a monotonically decreasing function and no minimum point exists. However, for $\epsilon > 0$ (i.e. $\alpha > 4$), the first term in (16) dominates as $P \rightarrow \infty$, indicating that there exists a minimum where the derivative of the curve changes its sign from negative to positive. We set $\epsilon > 0$ for what follows.

Differentiating $\mathcal{P}^{(c)}$ with respect to P we get

$$\begin{aligned} \frac{d\mathcal{P}^{(c)}}{dP} &= \frac{P^{-\frac{\epsilon+4}{\epsilon+2}} f_0^{-\frac{\lambda_u P^{\frac{2}{\epsilon+2}}}{A}}}{\epsilon+2} \times \\ & A(P\epsilon - 2P_s) f_0^{\frac{\lambda_u P^{\frac{2}{\epsilon+2}}}{A}} - 2\lambda_u f_0 P_d \log(f_0) P^{\frac{2}{\epsilon+2}} - 2A f_0 P_d. \end{aligned} \quad (17)$$

Given the fact that P is always positive, the derivative in (17) remains negative as P increases from 0 until it is sufficiently greater than $\frac{2P_s}{\epsilon}$, after which the derivative can change its sign to positive. This indicates that there exists a minimum value for $\mathcal{P}^{(c)}$ when

$$P^* > \frac{2P_s}{\epsilon}, \quad (18)$$

which concludes the proof. \blacksquare

For the case where BSs have no caching capabilities, the following result is derived.

Proposition 2. *Suppose that the BSs have no caching capabilities, thus $\mathcal{P}(P)$ is the objective function in (13). Then, for $\alpha = 4 + \epsilon$, $\epsilon > 0$, the optimal power allocation P^* that solves (13) is given by*

$$P^* = \frac{2(P_s + P_d)}{\epsilon}. \quad (19)$$

Proof: In the case without caching, by similar treatment as in the proof of Proposition 1, we write the expression for APC as

$$\mathcal{P} = \frac{A}{P^{\frac{2}{2+\epsilon}}} (P + P_s + P_d). \quad (20)$$

It can be noticed that \mathcal{P} is a monotonically decreasing function and has no minimum except when $\epsilon > 0$ (or $\alpha > 4$).

Differentiating \mathcal{P} with respect to P we get

$$\frac{d\mathcal{P}}{dP} = \frac{AP^{-\frac{\epsilon+4}{\epsilon+2}} (-2P_d - 2P_s + P\epsilon)}{\epsilon+2}. \quad (21)$$

By equating the derivative to zero, the optimum power

$$\frac{d\mathcal{P}}{dP} = 0 \Rightarrow P^* = \frac{2(P_s + P_d)}{\epsilon}. \quad (22)$$

It can easily be verified that the second derivative $\frac{d^2 \mathcal{P}}{dP^2} > 0$ for $P = P^*$. ■

A. Remarks

For a given finite value of transmit power P , from (16) and (20), we observe that, in all sensible cases,

$$\mathcal{P}^{(c)} < \mathcal{P}. \quad (23)$$

This indicates that BSs with caching capabilities always outperform those without caching. Additionally, $\mathcal{P}^{(c)}$ can be made smaller by increasing the catalog size f_0 in BSs. This is indeed intuitively correct, as shown in Fig. 2 where some realistic power values from [19] are considered.

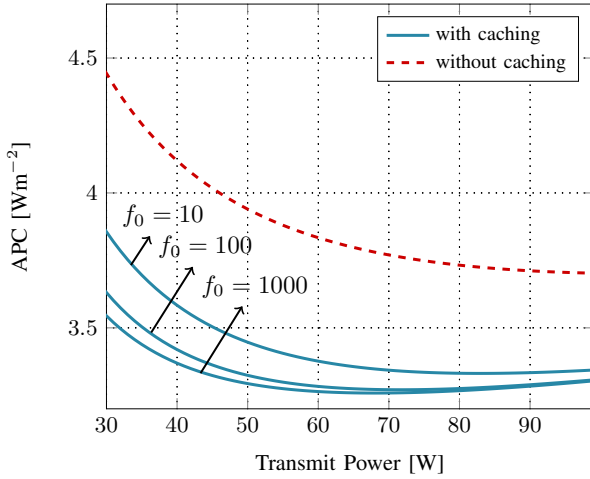


Figure 2: APC vs. Transmit power with and without caching for values: $P_s = 25\text{W}$, $P_d = 10\text{W}$, $\beta = 1$, and $\alpha = 4.75$ in (15) and (20).

Fig. 3 illustrates the variation of APC with respect to the transmit power in the cases with and without caching, for different values of pathloss exponent α . It can be noticed that the APC has a minimum for a certain power value only when $\alpha > 4$. However, APC can be significantly reduced with caching in all cases, and the performance gap between caching and no caching cases is increased for α increasing.

IV. ENERGY EFFICIENCY

Another key performance metric that should be studied is the energy key efficiency, which indicates the amount of utility (throughput) that is extracted out of a unit power invested for network operation. The standard QoS factor chosen as the utility in literature is the area spectral efficiency (ASE). In (24), we define the EE as the ratio between ASE and the total power spent at the BS, i.e.

$$\mathcal{E} = \frac{\lambda_b \log(1 + \gamma) P_{\text{cov}}(P, \lambda_b)}{P_{\text{tot}}}. \quad (24)$$

In order to define more precisely the EE metric for both cases, we use the expression for coverage probability established in

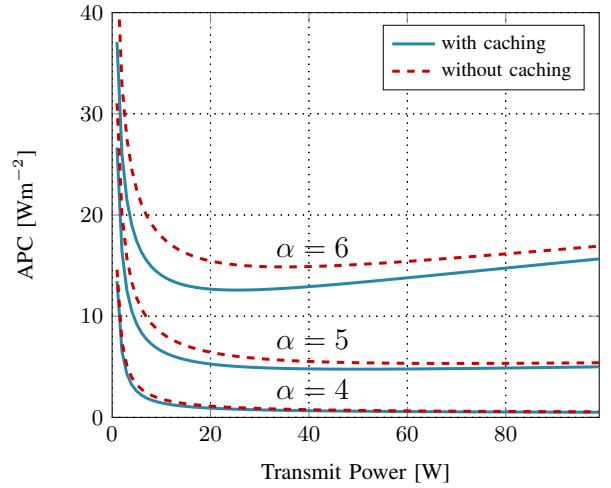


Figure 3: APC vs. Transmit power with and without caching for values: $P_s = 25\text{W}$, $P_d = 10\text{W}$, $f_0 = 10$, $A = 2$ in (15) and (20)

(5). With caching, the expression for EE in (24) therefore becomes

$$\mathcal{E}^{(c)}(P, \lambda_b) = \frac{\lambda_b \log(1 + \gamma) P_{\text{cov}}^{\text{NN}}(1 - \frac{A}{\lambda_b^{\frac{\alpha}{2}-1}} \frac{1}{P})}{P + P_s + P_d f_0^{1 - \frac{\lambda_u}{\lambda_b}}}. \quad (25)$$

In the case without caching, EE is given as

$$\mathcal{E}(P, \lambda_b) = \frac{\lambda_b \log(1 + \gamma) P_{\text{cov}}^{\text{NN}}(1 - \frac{A}{\lambda_b^{\frac{\alpha}{2}-1}} \frac{1}{P})}{P + P_s + P_d}. \quad (26)$$

The following results are given for the maximization of EE and the discussions are carried out afterwards.

Proposition 3. Suppose that the BSs have caching capabilities and let P_1^* denote the optimal power allocation that maximizes $\mathcal{E}^{(c)}(P, \lambda_b)$. Then

$$P_1^* = 1 + \sqrt{1 + P_s + P_d f_0^{1 - \frac{\lambda_u}{\lambda_b}}}. \quad (27)$$

Proof: We differentiate the expression (25) with respect to P and solve for the optimum power P_1^*

$$\begin{aligned} \frac{d\mathcal{E}^{(c)}}{dP} = 0 &\Rightarrow 2P - P^2 + P_s + P_d f_0^{1 - \frac{\lambda_u}{\lambda_b}} = 0 \\ &\Rightarrow P_1^* = 1 + \sqrt{1 + P_s + P_d f_0^{1 - \frac{\lambda_u}{\lambda_b}}}. \end{aligned} \quad (28)$$

It can be verified that the second derivative $\frac{d^2 \mathcal{E}^{(c)}}{dP^2}$ is negative for the positive solution of P_1^* . ■

Proposition 4. Suppose that the BSs have no caching capabilities and let P_2^* denote the optimal power allocation that maximizes $\mathcal{E}^{(c)}(P, \lambda_b)$. Then

$$P_2^* = 1 + \sqrt{1 + P_s + P_d}. \quad (29)$$

Proof: Similar to (28), we can show that the optimum power $P_2^* = 1 + \sqrt{1 + P_s + P_d}$. It can be verified that the

second derivative $\frac{d^2 \mathcal{E}}{dP^2}$ is negative for the positive solution of P_2^* . ■

A. Remarks

Based on the above results, we can make the following observations:

- 1) For given positive values of λ_b and P , $\mathcal{E}^{(c)}(P, \lambda_b)$ is always higher than $\mathcal{E}(P, \lambda_b)$ making it apparent that implementing BSs with caching capabilities is an energy-efficient solution.
- 2) For a fixed value of λ_b , $\mathcal{E}^{(c)}$ has a maximum at $P_1^* = 1 + \sqrt{1 + P_s + P_d f_0^{1-\frac{\lambda_u}{\lambda_b}}}$ and $\mathcal{E}(P, \lambda_b)$ has a maximum at $P_2^* = 1 + \sqrt{1 + P_s + P_d}$. Noting that $P_1^* < P_2^*$, we observe that the optimal EE may be attained at a smaller value of transmit power in the case of cache-enabled BSs.

In Fig. 4 we plot the variation of EE as a function of the transmit power P . It can be seen that EE can be significantly increased (in the case of caching) by increasing the size of the catalog in BSs, namely f_0 .

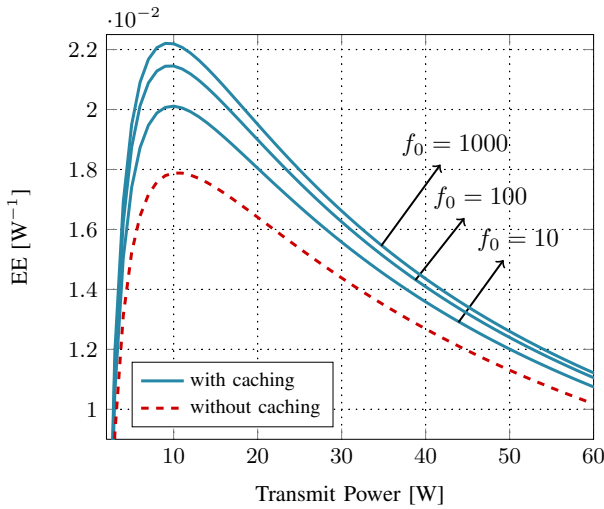


Figure 4: EE vs. transmit power with and without caching for values: $P_s = 25\text{W}$, $P_d = 10\text{W}$, $\alpha = 4.75$, $\beta = 1$, $P_{\text{cov}}^{\text{NN}} = 1$, $\lambda_b = 0.5$, $\lambda_u = 0.6$, and $\gamma = 2$ in (25) and (26).

V. CONCLUSIONS

In this work, we studied how incorporating caching capabilities at the BSs affects the energy consumption in wireless cellular networks. Adopting a detailed BS power model and modeling the BS locations according to a PPP, we derived expressions for the APC and the EE, which are further simplified in the low noise regime. A key observation of this work is that cache-enabled BSs can significantly decrease the APC and improve the EE as compared to traditional BSs. We also observed that the existence of an optimum power consumption point for the APC depends on the pathloss exponent.

The energy aspects and implications of caching in wireless cellular networks, especially for 5G systems, are of practical and timely interest and clearly require further investigation.

Future work may include heterogeneous network scenarios, including small cells, macro cells and WiFi access points deployment. Furthermore, storing the popular content requires accurate estimation of content popularity distribution, which cannot be easily performed in practice and may cost energy in terms of processing power. Therefore, rather than relying on this approach, randomized caching policies in a stochastic scenario [20] can be considered as a means to provide crisp insights on the energy efficiency benefits of caching in dense wireless networks.

REFERENCES

- [1] J. G. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *arXiv preprint: 1405.2957*, 2014.
- [2] E. Baştuğ, M. Bennis, and M. Debbah, "Living on the Edge: The role of proactive caching in 5G wireless networks," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 82–89, August 2014.
- [3] S. Göbbels, "Disruption tolerant networking by smart caching," *IEEE International Journal of Communication Systems*, pp. 569–595, April 2010.
- [4] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch, and G. Caire, "Femtocaching: Wireless video content delivery through distributed caching helpers," in *IEEE INFOCOM*, March 2012, pp. 1107–1115.
- [5] E. Baştuğ, J.-L. Guénégo, and M. Debbah, "Proactive small cell networks," in *20th International Conference on Telecommunications (ICT)*, Casablanca, Morocco, May 2013.
- [6] E. Baştuğ, M. Bennis, M. Kountouris, and M. Debbah, "Cache-enabled small cell networks: Modeling and tradeoffs," *EURASIP Journal on Wireless Communications and Networking*, Accepted (2015).
- [7] A. Altieri, P. Piantanida, L. R. Vega, and C. Galarza, "On fundamental trade-offs of device-to-device communications in large wireless networks," *arXiv preprint: 1405.2295*, 2014.
- [8] E. Baştuğ, M. Bennis, and M. Debbah, *Think Before Reacting: Proactive Caching in 5G Small Cell Networks*. Wiley, Submitted (2014).
- [9] R. Jain, C. So-In, and A.-k. Al Tamimi, "System-level modeling of IEEE 802.16E mobile WiMAX networks: key issues," *IEEE Wireless Commun.*, vol. 15, no. 5, pp. 73–79, October 2008.
- [10] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, November 2011.
- [11] C. S. Chen, V. M. Nguyen, and L. Thomas, "On small cell network deployment: A comparative study of random and grid topologies," in *IEEE VTC Fall*, September 2012.
- [12] C.-H. Lee, C.-Y. Shih, and Y.-S. Chen, "Stochastic geometry based models for modeling cellular networks in urban areas," *Wireless Networks*, vol. 19, no. 6, pp. 1063–1072, August 2013.
- [13] S. Mukherjee, *Analytical Modeling of Heterogeneous Cellular Networks*. Cambridge University Press, 2014.
- [14] H. S. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, "Modeling and analysis of K-tier downlink heterogeneous cellular networks," *IEEE J. Select. Areas Commun.*, vol. 30, no. 3, pp. 550–560, April 2012.
- [15] T. Samarasinghe, H. Inaltekin, and J. Evans, "Optimal SINR-based coverage in poisson cellular networks with power density constraints," in *IEEE VTC Fall*, September 2013.
- [16] B. Perabathini, M. Kountouris, M. Debbah, and A. Conte, "Optimal area power efficiency in cellular networks," in *IEEE Globecom Workshops (GC Wrokshops)*, Austin, Texas, December 2014. [Online]. Available: <http://goo.gl/ZP865y>
- [17] M. Z. Shafiq, L. Ji, A. X. Liu, and J. Wang, "Characterizing and modeling internet traffic dynamics of cellular devices," in *Proceedings of the ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems*. ACM, 2011, pp. 305–316.
- [18] M. E. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemporary physics*, vol. 46, no. 5, pp. 323–351, 2005.
- [19] M. Imran *et al.*, "Energy efficiency analysis of the reference systems, areas of improvements and target breakdown," ICT-EARTH deliverable, Tech. Rep., 2011. [Online]. Available: <http://www.ict-earth.eu>
- [20] B. Blaszczyk and A. Giovanidis, "Optimal geographic caching in cellular networks," *arXiv preprint: 1409.7626*, 2014.